# Armed Services Technical Information Agency



# AD 47733



Reproduced by

DOCUMENT SERVICE CENTER

KNOTT BUILDING, DAYTON, 2, OHIO

# UNCLASSIFIED

# OPTIMAL TEST LENGTH FOR MAXIMUM DIFFERENTIAL PREDICTION

Report of Research

Conducted for the Department of the Navy

Office of Naval Research Contract Nonr-477(08)

PAUL HORST
Professor of Psychology

November 1954

University of Washington
Seattle, Washington

# OPTIMAL TEST LENGTH FOR MAXIMUM DIFFERENTIAL PREDICTION[1]

## I THE PROBLEM

In (2) we have discussed the importance of techniques for predicting success differentially in each of a number of different activities from a single battery of predictors. It was assumed that intercorrelations for a large battery of predictor variables were available and also correlations between these predictors and a large number of criterion variables. The problem was to select from this larger battery of predictors that subset of specified size which would yield the maximum index of differential prediction for the criterion variables. The index of differential prediction efficiency was taken to be a simple function of the average of the variances for the predicted difference scores for all possible pairs of criterion variables. The larger this average variance the greater the differential prediction efficiency of the battery. It was shown that this index is equivalent to the difference between the average variance of the predicted criterion measures and the average of their covariances, assuming standard measures for both predictors and criteria and that the predicted criteria are the "least square" estimates. A method for selecting that subset of predictors of specified size which would yield the maximum index of differential prediction was presented.

The method referred to tacitly assumes that all predictors in the battery take the same amount of administration time, so that all subsets of the same size would also take the same amount of administration time. Usually this will not be the case. A more general approach to the problem might be to start with a given battery of predictor variables and inquire how the administration time for each predictor should be altered so that for a specified

over-all testing time the index of differential prediction efficiency will be a maximum. This approach would allow for increasing the length of an experimental battery as well as for decreasing it.

As a matter of fact for the case of a single criterion a method is already available (1) for determining the optimal distribution of testing time for a battery of predictors, assuming that intercorrelation, validity and reliability data are available for predictors of arbitrary lengths. It is the purpose of this article to present a modification and generalization of the method for the case of differential prediction involving a number of criterion variables.

In this presentation testing time is taken to be the time actually allotted the examinee for taking the test. A more complete analysis must also take into account the time for reading instructions, practice exercises, passing out and collecting papers, etc. The method will first be described and illustrated by a numerical example after which the mathematical rationale will be presented.

## II NUMERICAL EXAMPLE

The predictor variables used in this example are:

(1) Guilford-Zimmerman Aptitude Survey, Part I, Verbal Comprehension

(2) Guilford-Zimmerman Aptitude Survey, Part III, Numerical Operations

(3) Guilford-Zimmerman Aptitude Survey, Part VII, Mechanical Knowledge

(4) A. C. E. Psychological Examination, Quantitative Reasoning

(5) A. C. E. Psychological Examination, Linguistic Reasoning

(6) Cooperative English Test (Form OM), Usage

The matrix of test intercorrelations with reliabilities in the diagonal

is given in Table 1.

Table 1

R Matrix of Predictor Intercorrelations with Reliabilities Substituted for Unities in the Diagonal:

$$R = r - D_u$$

| | | 1 | 2 | 3 | 4 | 5 | 6 | Σ |
|---|---|---|---|---|---|---|---|---|
| 1 | G-Z 1 | .920 | .159 | .152 | .281 | .763 | .515 | 2.790 |
| 2 | G-Z 3 | .159 | .920 | .003 | .369 | .292 | .243 | 1.986 |
| 3 | G-Z 7 | .152 | .003 | .920 | .200 | .142 | -.150 | 1.267 |
| 4 | ACE-Q | .281 | .369 | .200 | .820 | .549 | .426 | 2.645 |
| 5 | ACE-L | .763 | .292 | .142 | .549 | .830 | .628 | 3.204 |
| 6 | English | .515 | .243 | -.150 | .426 | .628 | .860 | 2.522 |
| Σ | | 2.790 | 1.986 | 1.267 | 2.645 | 3.204 | 2.522 | 14.414 |

The criterion variables are grade point averages in each of ten college subjects. The matrix of validity coefficients is given in Table 2.

Table 2

The $r'_c$ Matrix of Validity Coefficients

| | 1<br>G-Z 1 | 2<br>G-Z 3 | 3<br>G-Z 7 | 4<br>ACE-Q | 5<br>ACE-L | 6<br>English | Σ |
|---|---|---|---|---|---|---|---|
| 1 Anthropology | .370 | .177 | .091 | .294 | .341 | .357 | 1.630 |
| 2 Chemistry | .317 | .274 | .016 | .309 | .364 | .399 | 1.679 |
| 3 Economics | .339 | .211 | .008 | .241 | .334 | .323 | 1.456 |
| 4 English | .526 | .247 | -.075 | .262 | .488 | .524 | 1.972 |
| 5 Foreign Lang. | .295 | .287 | -.156 | .200 | .232 | .426 | 1.284 |
| 6 Geology | .184 | .140 | .094 | .170 | .229 | .214 | 1.031 |
| 7 History | .379 | .169 | -.001 | .182 | .373 | .336 | 1.438 |
| 8 Mathematics | .287 | .348 | -.088 | .350 | .336 | .401 | 1.634 |
| 9 Psychology | .440 | .170 | .096 | .235 | .409 | .403 | 1.803 |
| 10 Zoology | .336 | .216 | .031 | .318 | .345 | .351 | 1.597 |
| Σ | 3.473 | 2.239 | .016 | 2.611 | 3.451 | 3.734 | 15.524 |
| Σ/10 | .347 | .224 | .002 | .261 | .345 | .373 | 1.552 |

The over-all testing time for the tests of arbitrary length is 142 minutes. We assume that this time is to be cut in half so that the over-all

testing time is 71 minutes. The problem is to determine the time to be allotted to each test so as to maximize the index of differential prediction efficiency.

The traditional assumptions are used here as in (1) with respect to the effect of test length on correlation and will not be repeated.

The method of solution for the new test lengths involves a series of successive approximations. For large numbers of predictor and criterion variables the solution may become very laborious. It is probable that the solution would be greatly expedited by the use of high speed computing equipment. Further research may yield more efficient computational procedures.

1. The first computational step is to calculate a matrix $\alpha_c'$ from the matrix $r_c'$ in Table 2. The elements $\alpha_c'$ of Table 3 are the corresponding elements of Table 2 with their column means subtracted. Hence the columns of Table 3 all add to zero.

Table 3

The $\alpha_c'$ Matrix: Validity Coefficients Expressed in Deviation Form for Each Test

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | .023 | -.047 | .089 | .033 | -.004 | -.016 |
| 2 | -.030 | .050 | .014 | .048 | .019 | .026 |
| 3 | -.008 | -.013 | .006 | -.020 | -.011 | -.050 |
| 4 | .179 | .023 | -.077 | .001 | .143 | .151 |
| 5 | -.052 | .063 | -.158 | -.061 | -.113 | .053 |
| 6 | -.163 | -.084 | .092 | -.091 | -.116 | -.159 |
| 7 | .032 | -.055 | -.003 | -.079 | .028 | -.037 |
| 8 | -.060 | .124 | -.090 | .089 | -.009 | .028 |
| 9 | .093 | -.054 | .094 | .024 | .064 | .030 |
| 10 | -.011 | -.008 | .029 | .057 | .000 | -.022 |
| Ck | .003 | -.001 | -.004 | .001 | .001 | .004 |
| Σ | .003 | -.001 | -.004 | .001 | .001 | .004 |

2. The next step is to compute the elements for a diagonal matrix, $\Delta$. The i'th element is the product of the original length of test i multiplied by one minus its reliability. The elements for $\Delta$ are given in row 4, labeled $1'\Delta$ of Table 4. For the first element we have $\Delta_1 = 25(1.00 - .92) = 2.00$.

Table 4

Computation of $1'D_{b_i}^{-1}$ and $1'\Delta D_{b_i}^{-1}$

First approximation: $1'D_{b_1} = \left( \frac{T_1}{1'D_a 1} \right) 1'D_a$

| | | 1 | 2 | 3 | 4 | 5 | 6 | Ck | $\Sigma$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $1'D_a$ | 25.0 | 9.0 | 30.0 | 23.0 | 15.0 | 40.0 | | 142.0 |
| 2 | $1'D_{b_1} = .5\ 1'D_a$ | 12.5 | 4.5 | 15.0 | 11.5 | 7.5 | 20.0 | 71.0 | 71.0 |
| 3 | $1'D_{b_1}^{-1}$ | .080 | .222 | .0667 | .087 | .1333 | .050 | | |
| 4 | $1'\Delta$ | 2.000 | .720 | 2.400 | 4.140 | 2.550 | 5.600 | | 17.410 |
| 5 | $1'\Delta D_{b_1}^{-1}$ | .160 | .160 | .160 | .360 | .340 | .280 | | 1.460 |

3. A first approximation is now required for the altered test lengths. We assume the new test lengths to be proportional to the original test lengths. Therefore, as a first approximation to the new test lengths we take one half the original test lengths. Row 1 in Table 4 gives the original test lengths. Row 2 of the same table is one half the first.

4. Calculate the reciprocals of the $D_b$ elements. These are given in row 3 of Table 4.

5. Calculate the product of each $\Delta$ value in row 4 of Table 4 by the corresponding value immediately above it. The ratios are entered in row 5 of Table 4. For example, the first value is $.160 = 2.000 \times .080$.

6. Next the elements calculated in step 5 are added to the corresponding diagonal elements of Table 1, and the table is copied into the upper left quadrant of Table 5. The first diagonal element is $1.080 = .160 + .920$.

Note that the elements below the diagonal are not copied in. The upper right section of Table 5 is $\alpha_c$ , the transpose of Table 3.

7. We next calculate a matrix $L_1$ by premultiplying the matrix $\alpha_c$ by the inverse of the matrix in the upper left quadrant of Table 5. The computations for the forward solution are given in the two lower quadrants of Table 5 and in Table 6. The back solution is given in Table 7. The procedure for multiplying a matrix by the inverse of a symmetric matrix is outlined in (3).

8. The second approximation to the new test lengths is computed in the lower section of Table 7 as follows:

Row a consists of the sum of squares of column elements of the $L_1$ matrix. For example, the first element in row a, namely, .0626, is the sum of squares of the first ten elements in column 1 of Table 7.

Row b is copied from row 4 of Table 4.

Row c consists of the products of corresponding elements in the two preceding lines. For example, $.1251 = .0626 \times 2.00$ .

Row d consists of the square roots of corresponding entries in the preceding line. For example, $.3537 = \sqrt{.1251}$ . The computations to the right of this line and designated s are obtained by dividing the over-all new testing time, 71 minutes, by 1.8823, the sum of the elements in the row. This gives $s = 37.7198$.

Row e is a check row. Each element in the second line above it is divided by the element immediately above. Thus $\frac{.1251}{.3537} = .3537$ .

Row f is obtained by multiplying each element in row d by s. For example, the first element is $13.3415 = .3537 \times 37.7198$ . This line gives the second approximation to the new test lengths.

Row g is obtained by dividing each element in the preceding row into the corresponding value in row b. For example, the first value is $.150 = \frac{2.00}{13.3415}$ .

Table 5

Computation of $(R+\Delta D_{b_1}^{-1})^{-1}\,\alpha_c = L_1$ Forward Solution

| | | 1A | 2A | 3A | 4A | 5A | 6A | 1B | 2B | 3B | 4B | 5B | 6B | 7B | 8B | 9B | 10B | Check | Σ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1A | 1.080 | .159 | .152 | .281 | .763 | .515 | .023 | -.030 | -.008 | .179 | -.052 | -.163 | .032 | -.060 | .093 | -.011 | 2.953 | |
| | 2A | | 1.080 | .003 | .369 | .292 | .243 | -.047 | .050 | -.013 | .023 | .063 | -.084 | -.055 | .124 | -.054 | -.008 | 2.145 | |
| | 3A | | | 1.080 | .200 | .142 | -.150 | .089 | .014 | .006 | -.077 | -.158 | .092 | -.003 | -.090 | .094 | .029 | 1.423 | |
| | 4A | | | | 1.180 | .549 | .426 | .033 | .048 | -.020 | .001 | -.061 | -.091 | -.079 | .089 | .024 | .057 | 3.006 | |
| | 5A | | | | | 1.170 | .628 | -.004 | .019 | -.011 | .143 | -.113 | -.116 | .028 | -.009 | .064 | .000 | 3.545 | |
| | 6A | | | | | | 1.140 | -.016 | .026 | -.050 | .151 | .053 | -.159 | -.037 | .028 | .030 | -.022 | 2.806 | |
| | | 2.950 | 2.146 | 1.427 | 3.005 | 3.544 | 2.802 | .078 | .127 | -.096 | .420 | -.268 | -.521 | -.114 | .082 | .251 | .045 | 15.878 | |
| .9259 | 1 | 1.080 | .159 | .152 | .281 | .763 | .515 | .023 | -.030 | -.008 | .179 | -.052 | -.163 | .032 | -.060 | .093 | -.011 | 2.953 | 2.953 |
| .9461 | 2 | | 1.057 | -.019 | .328 | .180 | .167 | -.050 | .054 | -.012 | -.003 | .071 | -.061 | -.060 | .133 | -.068 | -.006 | 1.711 | 1.711 |
| .9452 | 3 | | | 1.058 | .166 | .038 | -.220 | .085 | .019 | .007 | -.102 | -.149 | .114 | -.009 | -.079 | .080 | .030 | 1.037 | 1.038 |
| 1.0215 | 4 | | | | .979 | .289 | .275 | .029 | .036 | -.015 | -.029 | -.046 | -.048 | -.067 | .076 | .008 | .057 | 1.545 | 1.544 |
| 1.9455 | 5 | | | | | .514 | .163 | -.023 | .020 | .001 | .029 | -.069 | .020 | .036 | -.009 | .005 | -.009 | .676 | .678 |
| 1.4430 | 6 | | | | | | .693 | -.002 | .019 | -.039 | .044 | .070 | -.041 | -.037 | .001 | .009 | -.023 | .694 | .694 |

Table 6

Computation of $(R+\Delta D_{b_1}^{-1})^{-1} \alpha_c = L_1$ continued

| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Check | Σ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | | | | | | | | | | | | | | | | | |
| 2 | | 1.000 | | | | | | | | | | | | | | | | |
| 3 | | | 1.000 | | | | | | | | | | | | | | | |
| 4 | | | | 1.000 | | | | | | | | | | | | | | |
| 5 | | | | | 1.000 | | | | | | | | | | | | | |
| 6 | | | | | | 1.000 | | | | | | | | | | | | |
| 1 | -1.000 | -.147 | -.141 | -.260 | -.796 | -.477 | -.021 | .028 | .007 | -.166 | .048 | .151 | -.030 | .056 | -.086 | .010 | -2.734 | -2.734 |
| 2 | | -1.000 | .018 | -.310 | -.170 | -.158 | .047 | -.051 | .011 | .003 | -.067 | .057 | -.057 | -.126 | .064 | .006 | -1.619 | -1.619 |
| 3 | | | -1.000 | -.157 | -.036 | .208 | -.080 | -.018 | -.007 | .096 | .141 | -.108 | .009 | .075 | -.076 | -.028 | - .981 | - .981 |
| 4 | | | | -1.000 | -.295 | -.281 | -.030 | -.037 | .015 | .030 | .047 | .049 | .068 | -.077 | -.008 | -.058 | -1.577 | -1.577 |
| 5 | | | | | -1.000 | -.317 | .045 | -.039 | -.002 | -.056 | .134 | -.039 | -.070 | .018 | -.010 | .018 | -1.319 | -1.318 |
| 6 | | | | | | -1.000 | .003 | -.027 | .056 | -.063 | -.101 | .059 | .053 | -.001 | -.013 | .033 | -1.001 | -1.001 |

Table 7

Computation of $(R+\Delta D^{-1}_{b_1})^{-1}\alpha_c = L_1$ Back Solution

$L_1'$ Matrix

| | | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Check | | $L_1$ 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | .039 | -.051 | .074 | .044 | -.044 | -.003 | -1 | | | | | | | | | | 0 | .001 | .255 |
| 2 | | -.075 | .035 | .019 | .021 | .030 | .027 | | -1 | | | | | | | | | 0 | -.001 | .207 |
| 3 | | .003 | -.004 | -.005 | -.005 | .020 | -.056 | | | -1 | | | | | | | | 0 | .003 | .098 |
| 4 | | .136 | -.002 | -.075 | -.058 | .036 | .063 | | | | -1 | | | | | | | 0 | .000 | .370 |
| 5 | | .031 | .085 | -.110 | -.026 | -.166 | .101 | | | | | -1 | | | | | | 0 | .001 | .519 |
| 6 | | -.159 | -.040 | .101 | -.050 | .058 | -.059 | | | | | | -1 | | | | | 0 | .000 | .467 |
| 7 | | .022 | -.039 | -.011 | -.079 | .087 | -.053 | | | | | | | -1 | | | | 0 | .002 | .291 |
| 8 | | -.068 | .102 | -.087 | .082 | -.018 | .001 | | | | | | | | -1 | | | 0 | .002 | .358 |
| 9 | | .074 | -.067 | .078 | .003 | .006 | .013 | | | | | | | | | -1 | | 0 | .002 | .241 |
| 10 | | -.005 | -.021 | .010 | .070 | -.008 | -.033 | | | | | | | | | | -1 | 0 | -.001 | .147 |
| a | $1'D_{L_1L_1'}$ | .0626 | .0295 | .0477 | .0268 | .0434 | .0256 | | Ck | | | Σ | | | | | | | | |
| b | $1'\Delta$ | 2.00 | .72 | 2.40 | 4.14 | 2.55 | 5.60 | | 17.41 | | | 17.41 | | | | | | | | |
| c | $1'D_{L_1L_1'}\Delta$ | .1251 | .0214 | .1144 | .1110 | .1108 | .1433 | | | | | | | | | | | | | |
| d | $1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}$ | .3537 | .1458 | .3382 | .3332 | .3328 | .3786 | | | | | | 1.8823 | | | $s = \frac{T_1}{1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}1} =$ | | | | |
| e | Ck: $1'D_{L_1L_1'}\Delta(D_{L_1L_1'}\Delta)^{-\frac{1}{2}}$ | .3537 | .1458 | .3362 | .3332 | .3329 | .3786 | | | | | | | | | | | | | |
| f | $1'D_{b_2} = 1'(D_{L_1L_1'}\Delta)^{\frac{1}{2}}s$ | 13.3415 | 5.4995 | 12.7568 | 12.5682 | 12.5531 | 14.2807 | | | | | | 71.0000 | | | $\frac{71}{1.8823} = 37.7198$ | | | | |
| g | $1'\Delta D^{-1}_{b_2}$ | .150 | .131 | .138 | .329 | .203 | .392 | | | | | | 1.393 | | | | | | | |
| h | Ck: $1'D_{b_2}\Delta D^{-1}_{b_2}$ | 2.001 | .720 | 2.398 | 4.135 | 2.548 | 5.598 | | | | | | | | | | | | | |
| i | $1'D_{r_{ii}} + 1'\Delta D^{-1}_{b_2}$ | 1.070 | 1.051 | 1.108 | 1.149 | 1.033 | 1.252 | | 6.663 | | | | 6.663 = 1' Diag. of $(R+\Delta D^{-1}_{b_2})$ | | | | | | | |

Row h is a check on the preceding row. Each element in this row is multiplied by the corresponding element in row f to give row b. For example, the first element is 2.001 = .150 x 13.3415.

Row i is obtained by adding the elements in row g to the corresponding reliabilities. For example, the first element is 1.070 = .150 + .920.

9. A new $L_2$ matrix is now computed by repeating steps 6 and 7 and using the elements of row i of Table 7 in the diagonal positions of Table 1. The new $L_2$ matrix is given in transposed form in Table 8, rows 1 through 10.

10. Step 8 is repeated in rows a through i of Table 8. Row f of this table gives a third approximation to the altered test lengths.

Steps 6, 7 and 8 are repeated to get succeeding approximations to the test lengths. The calculations were carried to 5 successive approximations for the new test lengths, not counting the first. These are summarized in Table 9. As will be seen, the iterations have not completely stabilized. However, for practical purposes, the approximation is doubtless adequate.

11. To compute the successive indices of differential prediction efficiency, $\phi_c$ , we proceed as follows:

(a) For the index corresponding to the first approximation to the new test length multiply each element in the $L_1$ matrix in Table 7 by the corresponding element of Table 3 and sum the products. This is the first entry, .227, in the $\phi$ column at the right of Table 9.

(b) To get $\phi_2$ follow the same procedure except use the $L_2$ matrix in Table 8 instead of $L_1$ in Table 7.

(c) In the same way calculate subsequent $\phi$'s by using the elements in the corresponding L matrix and the elements in Table 3.

Table 8

Computation of $(R+\Delta D^{-1}_{b_2})^{-1}\,\alpha_c = L_2$ Back Solution

$L_2'$ Matrix

| | | 1 | 2 | 3 | 4 | 5 | 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | -.051 | -.053 | .072 | .052 | -.065 | .000 | | |
| 2 | | -.083 | .036 | .017 | .018 | .045 | .022 | | |
| 3 | | .003 | -.005 | -.003 | -.009 | .024 | -.049 | | |
| 4 | | .132 | -.003 | -.076 | -.062 | .053 | .052 | | |
| 5 | | .070 | .093 | -.108 | -.009 | -.229 | .100 | | |
| 6 | | -.176 | -.043 | .100 | -.057 | .081 | -.056 | | |
| 7 | | .002 | -.041 | -.009 | -.092 | .119 | -.052 | | |
| 8 | | -.066 | .105 | -.085 | .086 | -.025 | .003 | | |
| 9 | | .076 | -.069 | .075 | .005 | .007 | .010 | | |
| 10 | | -.005 | -.022 | .012 | .072 | -.012 | -.028 | | |
| a | $1'D_{L_2L_2'}$ | .0730 | .0326 | .0460 | .0314 | .0836 | .0223 | | |
| b | $1'\Delta$ | 2.0 | .72 | 2.40 | 4.14 | 2.55 | 5.60 | | |
| c | $1'D_{L_2L_2'}\Delta$ | .1459 | .0235 | .1104 | .1298 | .2132 | .1250 | | |
| d | $1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}$ | .3820 | .1532 | .3323 | .3603 | .4618 | .3536 | Ck | Σ 2.0432 |
| e | Ck: $1'D_{L_2L_2'}\Delta(D_{L_2L_2'}\Delta)^{-\frac{1}{2}}$ | .3820 | .1532 | .3322 | .3602 | .4617 | .3535 | | |
| f | $1'D_{b_3} = 1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}s$ | 13.2743 | 5.3236 | 11.5472 | 12.5202 | 16.0473 | 12.2874 | 71.000 | 71.0000 |
| g | $1'\Delta D^{-1}_{b_3}$ | .151 | .135 | .208 | .331 | .159 | .456 | | 1.440 |
| h | Ck: $1'D_{b_3}\Delta D^{-1}_{b_3}$ | 2.004 | .720 | 2.402 | 4.144 | 2.552 | 5.603 | | |
| i | $1'D_{r_{ii}} + 1'\Delta D^{-1}_{b_3}$ | 1.071 | 1.055 | 1.128 | 1.151 | .989 | 1.316 | 6.710 | 6.710 = $1'$ Diag. of $(R+D^{-1}_{b_3})$ |

$$s = \frac{T_1}{1'(D_{L_2L_2'}\Delta)^{\frac{1}{2}}1} = \frac{71}{2.0432} = 34.7494$$

Table 9

Successive Approximations to $1'D_b$ , for $T_1 = \frac{1}{2}T_0 = \frac{142}{2} = 71$

| Approx'n | | 1 | 2 | 3 | 4 | 5 | 6 | Σ | Value of Ø for Successive Values of L | |
|---|---|---|---|---|---|---|---|---|---|---|
| $(.5)1'D_a$: | 1 | 12.50 | 4.50 | 15.00 | 11.50 | 7.50 | 20.00 | 71.00 | | |
| | 2 | 13.34 | 5.50 | 12.76 | 12.57 | 12.55 | 14.28 | 71.00 | $L_1$ | .227 |
| | 3 | 13.27 | 5.32 | 11.55 | 12.52 | 16.05 | 12.29 | 71.00 | $L_2$ | .234 |
| | 4 | 13.23 | 5.20 | 10.98 | 12.47 | 17.62 | 11.51 | 71.01 | $L_3$ | .235 |
| | 5 | 13.31 | 5.15 | 10.76 | 12.46 | 18.13 | 11.19 | 71.00 | $L_4$ | .236 |
| | 6 | 13.35 | 5.12 | 10.70 | 12.46 | 18.37 | 11.00 | 71.00 | $L_5$ | .237 |

It will be noted that Ø does not increase much in this particular illustration. It goes from .227, taking the test lengths as one half their original length, to .237 as they approach optimal length. This is an increase of less than 5 per cent even though several of the test lengths are changed greatly. For example, test 5 increases from 7 to 18 minutes while test 6 reduces from 20 to 11 minutes.

The technique was applied to the same data assuming that the total administration time was to be the same as in the original administration, namely, 142 minutes and also assuming it was to be doubled to 284 minutes. Only three approximations to the optimal test lengths were calculated for each of these two conditions. Tables 10 and 11 summarize the results for the two conditions respectively. The last column in each table shows the index of differential prediction efficiency, Ø, corresponding to each approximation to optimal test lengths. In both cases the improvement of Ø as the tests approach optimal length is appreciably greater than for the case of one half the original testing time. This rate of improvement is greatest for double testing time. As can be seen from the right hand column of Table 11

it goes from .305 to .337 which is approximately a 10 per cent increase. Further research is needed to determine the sensitivity of Ø to alterations in relative testing time for each of the tests and to variation in total testing time.

Table 10

Successive Approximations to $1'D_b$, for $T_1 = T_0 = 142$

| Approx'n | | 1 | 2 | 3 | 4 | 5 | 6 | Σ | Value of Ø for Successive Values of L | |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) $1'D_a$: | 1 | 25.00 | 9.00 | 30.00 | 23.00 | 15.00 | 40.00 | 142.00 | | |
| | 2 | 25.38 | 5.85 | 20.04 | 25.55 | 33.62 | 28.56 | 142.00 | $L_1$ | .265 |
| | 3 | 25.83 | 7.60 | 16.02 | 25.41 | 42.43 | 24.73 | 142.02 | $L_2$ | .282 |
| | 4 | 26.48 | 7.30 | 14.72 | 25.22 | 44.54 | 23.73 | 141.99 | $L_3$ | .283 |

Table 11

Successive Approximations to $1'D_b$, for $T_1 = 2T_0 = 2(142) = 284$

| Approx'n | | 1 | 2 | 3 | 4 | 5 | 6 | Σ | Value of Ø for Successive Value of L | |
|---|---|---|---|---|---|---|---|---|---|---|
| (2) $1'D_a$: | 1 | 50.00 | 18.00 | 60.00 | 46.00 | 30.00 | 80.00 | 284.00 | | |
| | 2 | 51.51 | 13.62 | 30.14 | 51.31 | 83.57 | 53.84 | 284.00 | $L_1$ | .305 |
| | 3 | 54.84 | 10.84 | 22.51 | 51.17 | 97.72 | 46.91 | 284.00 | $L_2$ | .334 |
| | 4 | 56.32 | 10.53 | 21.24 | 50.94 | 99.64 | 45.34 | 284.01 | $L_3$ | .337 |

## III MATHEMATICAL DERIVATION

In (1) a procedure is developed for altering test lengths in a battery to give maximum multiple correlation with a single criterion. The development of this procedure will be reviewed and the procedure will be extended to the problem of differential prediction.

Let

$M$ be the number of cases,

$n$ be the number of predictors,

$Z$ be an $(M \times n)$ matrix of test scores in a battery of altered lengths with the elements of $Z$ of the form $\frac{z_{ij} - \bar{z}_j}{\sqrt{M}\ \sigma_{z_j}}$ ,

$W$ be an $(M \times 1)$ vector of criterion scores with elements of the form $\frac{w - \bar{w}}{\sqrt{M}\ \sigma_w}$ ;

$B$ be an $(M \times 1)$ vector of regression coefficients for estimating $W$ from $Z$,

$r$ be an $(n \times n)$ matrix of intercorrelations of tests of original lengths,

$\rho$ be an $(n \times n)$ matrix of intercorrelations of tests of altered lengths,

$r_c$ be an $(n \times 1)$ vector of validity coefficients for the tests of original lengths,

$\rho_c$ be an $(n \times 1)$ vector of validity coefficients for the tests of altered lengths,

$D_a$ be an $(n \times n)$ diagonal matrix of original test lengths,

$D_b$ be an $(n \times n)$ diagonal matrix of altered test lengths,

$D_e = D_b D_a^{-1}$ be the ratio of altered to original test lengths,

$D_{r_{ii}}$ be the $(n \times n)$ diagonal matrix of reliability coefficients for the tests of original lengths.

Let $\delta = \left[ I + (D_e - I) D_{r_{ii}} \right] D_e^{-1}$ . (1)

Let $\varepsilon = (ZB - W)$ . (2)

We wish to minimize $\varepsilon'\varepsilon$ with the constraining condition $1'D_b 1 = T$ where T is the ratio of the new total testing time to the original total testing time, and 1 is a column vector of all unit elements.

To obtain $\varepsilon'\varepsilon$ minimum under this condition

Let $\psi = \varepsilon'\varepsilon + \lambda 1' D_b 1$ (3)

where $\lambda$ is a Lagrangian multiplier.

From (2)

$$\psi = (B'Z'ZB - B'Z'W - W'ZB + W'W) + \lambda 1'D_b 1 \quad . \tag{4}$$

From the definitions above

$$Z'Z = \rho \quad , \tag{5}$$

$$Z'W = \rho_c \quad , \tag{6}$$

$$W'W = 1 \quad . \tag{7}$$

Substituting (5), (6), and (7) in (4)

$$\psi = B'\rho B - B'\rho_c - \rho_c' B + 1 + \lambda 1'D_b 1 \quad . \tag{8}$$

In (1) it is shown that

$$\rho_c = \delta^{-\frac{1}{2}} r_c \tag{9}$$

and

$$\rho = \delta^{-\frac{1}{2}}(r - D_u + D_u D_a D_b^{-1})\delta^{-\frac{1}{2}} \tag{10}$$

where we define $D_u = I - D_{r_{ii}}$ , a diagonal matrix of test unreliability coefficients.

Let $B = \delta^{\frac{1}{2}}\beta$ . (11)

Substituting (9), (10), and (11) in (8)

$$\psi = \beta'(r - D_u + D_u D_a D_b^{-1})\beta - \beta' r_c - r_c'\beta + 1 + \lambda 1'D_b 1 \quad . \tag{12}$$

The unknowns on the right hand side of (12) are $\beta$, $D_b$ and $\lambda$ .

Differentiating (12) with respect to $\beta'$ and equating the resulting expression to zero to get an extremum

$$\frac{\partial \psi}{\partial \beta'} = r_c - (r - D_u + D_u D_a D_b^{-1})\beta = 0$$

or $\quad \beta = (r - D_u + D_u D_a D_b^{-1})^{-1} r_c \qquad (13)$

Differentiating $\psi$ with respect to the scalars, $b_i$, $(i = 1, 2, \ldots, n)$ and equating the n resulting expressions to zero

$$\frac{\partial \psi}{\partial b_i} = \lambda - \frac{\beta_i^2 (u_i a_i)}{b_i^2} = 0 \qquad (14)$$

or

$$b_i = \frac{\beta_i (u_i a_i)^{\frac{1}{2}}}{\lambda^{\frac{1}{2}}} \qquad (15)$$

Summing these n equations

$$\Sigma b_i = \frac{1}{\lambda^{\frac{1}{2}}} \Sigma \beta_i (u_i a_i)^{\frac{1}{2}}$$

Thus in matrix notation we obtain

$$\lambda^{\frac{1}{2}} = \frac{1'(D_u D_a)^{\frac{1}{2}} \beta}{1' D_b 1} \qquad (16)$$

Substituting for $\lambda^{\frac{1}{2}}$ in (15) and collecting these n expressions as the diagonal matrix $D_b$ we obtain

$$D_b = D_\beta (D_u D_a)^{\frac{1}{2}} \frac{1' D_b 1}{1'(D_u D_a)^{\frac{1}{2}} \beta} \qquad (17)$$

where $D_\beta$ is a diagonal matrix with the $\beta_i$ as diagonal elements.

In (1) it is shown that

$$\beta = \left( r - D_u + \frac{\left[ (D_u D_a)^{\frac{1}{2}} 1 \right] \left[ 1' (D_u D_a)^{\frac{1}{2}} \right]}{1' D_b 1} \right)^{-1} r_c \qquad (18)$$

Using (18) in (17) we can therefore solve for $D_b$, the new test lengths. The new multiple correlation is given by

$$R_b^2 = \beta' r_c \quad . \tag{19}$$

Next we extend the procedure to the case of differential prediction. Consider the following additions to the definitions given above.

Let

- N be the number of criteria,
- W be an (M x N) matrix of criterion scores whose elements are deviate scores of the form $\frac{w_{ij} - \bar{w}_j}{\sqrt{M}\,\sigma_{w_j}}$ ,
- H be an (M x $N^2$) matrix consisting of difference vectors for all possible pairs of criterion vectors i and j, including i = j,
- B be an (n x $N^2$) matrix of "least square" regression vectors for estimating H from X,
- $r_c$ be the (n x N) matrix of validity coefficients with the tests of original lengths,
- $\rho_c$ be the (n x N) matrix of validity coefficients with the tests of altered lengths.

From the differential prediction procedure (2) we have

$$\phi = 1'D_c1 - \frac{1'C1}{N} \text{ , the index of differential prediction efficiency,} \tag{20}$$

where $C = r_c' \, r^{-1} \, r_c$ (21)

and $D_c$ is a diagonal matrix of the diagonals of C.

Let $E = ZB - H$ (22)

and $F_i = e_i 1' - I$ (23)

where $e_i$ is a column vector of all zero elements except the i'th which is unity.

Let $G' = (F_1, F_2, \ldots, F_N)$ . (24)

Thus we have

$$H = WG' \tag{25}$$

and

$$E = ZB - WG' \quad . \tag{26}$$

From (23), (24) and (26) postmultiplied by G and divided by 2N we obtain

$$\varepsilon = E\frac{G}{2N} = Z\frac{BG}{2N} - W\left(I - \frac{11'}{N}\right) \tag{27}$$

since $G'G = 2N\left(I - \frac{11'}{N}\right)$ .

Let $B\frac{G}{2N} = J$ . (28)

Let $W\left(I - \frac{11'}{N}\right) = t$ . (29)

Then

$$\varepsilon = ZJ - t \quad . \tag{30}$$

We wish to minimize the trace of $\varepsilon'\varepsilon$ with the constraining condition

$$1'D_b1 = T \ .$$

Let $\psi = \operatorname{tr} \varepsilon'\varepsilon + \lambda 1'D_b1$ . (31)

Let $\gamma_c = \rho_c\left(I - \frac{11'}{N}\right)$ . (32)

Substituting (5), (6), (9), (10), (28), (29), (30), and (32) in (31)

we obtain

$$\psi = \operatorname{tr}\left[J'\delta^{-\frac{1}{2}}(r - D_u + D_uD_aD_b^{-1})\delta^{-\frac{1}{2}}J - J'\gamma_c - \gamma_c'J + t't\right] + \lambda 1'D_b1 \quad . \tag{33}$$

Let $\delta^{-\frac{1}{2}}J = L$ , (34)

Let $\alpha_c = r_c\left(I - \frac{11'}{N}\right) = \delta^{\frac{1}{2}}\gamma_c$ . (35)

Let $R = r - D_u$ . (36)

Let $\Delta = D_uD_a$ . (37)

Substituting (34), (35), (36) and (37) in (33) we obtain

$$\psi = \operatorname{tr}\left[L'(R + \Delta D_b^{-1})L - L'\alpha_c - \alpha_c'L + t't\right] + \lambda 1'D_b1 \quad . \tag{38}$$

Differentiating (38) with respect to row vectors of $L'$ and equating the results to 0 we obtain

$$\frac{\partial \psi}{\partial L'} = \alpha_c - (R + \Delta D_b^{-1})L = 0$$

or

$$\alpha_c = (R + \Delta D_b^{-1})L \quad . \tag{39}$$

Differentiating (37) with respect to $D_b$ and equating the results to 0 we obtain

$$\frac{\partial \psi}{\partial D_b} = \lambda I - D_{LL'}\Delta D_b^{-2} = 0 \tag{40}$$

where $D_{LL'}$ is a diagonal matrix whose non-zero elements are the diagonal elements of $LL'$ .

Hence

$$D_b = \frac{1}{\lambda^{\frac{1}{2}}} (D_{LL'}\Delta)^{\frac{1}{2}} \quad . \tag{41}$$

It can be shown that

$$\lambda^{\frac{1}{2}} = \frac{1'(D_{LL'}\Delta)^{\frac{1}{2}} 1}{1' D_b 1} = \frac{1'(D_{LL'}\Delta)^{\frac{1}{2}} 1}{T} \quad . \tag{42}$$

Substituting (42) in (41)

$$D_b = (D_{LL'}\Delta)^{\frac{1}{2}} \frac{T}{1'(D_{LL'}\Delta)^{\frac{1}{2}} 1} \quad . \tag{43}$$

From (39)

$$L = (R + \Delta D_b^{-1})^{-1} \alpha_c \quad . \tag{44}$$

Let

$$L_i = (R + \Delta D_{b_i}^{-1})^{-1} \alpha_c \tag{45}$$

where

$$D_{b_1} = \frac{T}{1' D_a 1} D_a \quad . \tag{46}$$

and

$$D_{b_{i+1}} = \frac{(D_{L_i L_i'} \Delta)^{\frac{1}{2}} T}{1'(D_{L_i L_i'} \Delta)^{\frac{1}{2}} 1} . \tag{47}$$

Using (45), (46) and (47) as a basis of successive approximations to $L_i$ and $D_{b_i}$ continue until $D_{b_i}$ stabilizes satisfactorily.

Now the regression vectors for the optimal test lengths will be given by

$$Y = \rho^{-1}\rho_c . \tag{48}$$

From (9), (10), (36), (37) and (48)

$$Y = \delta^{\frac{1}{2}}(R + \Delta D_b^{-1})^{-1} r_c . \tag{49}$$

But from (35) and (39)

$$L = (R + \Delta D_b^{-1})^{-1} r_c \left(I - \frac{11'}{N}\right) . \tag{50}$$

From (50)

$$\delta^{\frac{1}{2}}L = \delta^{\frac{1}{2}}(R + \Delta D_b^{-1})^{-1} r_c - \delta^{\frac{1}{2}}(R + \Delta D_b^{-1})^{-1} r_c \frac{11'}{N} . \tag{51}$$

From (49 and (51)

$$Y = \delta^{\frac{1}{2}}\left[L + (R + \Delta D_b^{-1})^{-1} r_c \frac{11'}{N}\right] . \tag{52}$$

Furthermore the index of differential prediction efficiency "$\phi$" as defined in (2) can be shown to be

$$\phi = \mathrm{tr}\ L' \alpha_c . \tag{53}$$

The procedures outlined in Section II may be related to the above mathematical development as follows:

Table 1 is given by equation (36).

Step 1 is based on equation (35).

Step 2 is based on equation (37).

Step 3 is based on equation (46).

Step 4 consists of calculating $D_{b_1}^{-1}$ from $D_{b_1}$ .

Step 5 consists of calculating $\Delta D_{b_1}^{-1}$ .

Step 6 consists of calculating the parenthesis on the right side of equation (45) for $i = 1$.

Step 7 consists of calculating $L_1$ from equation (45).

Step 8 consists of calculating $D_{b_2}$ from equation (47).

Step 9 consists of calculating an $L_2$ matrix from equation (45).

Step 10 uses equation (47) to calculate $D_{b_3}$.

In general steps 6 and 7 are repeated for successive values of i in equation (45) and step 8 is repeated for successive values of i in equation (47).

Step 11 uses equation (53) to get successive values of $\phi$.

## REFERENCES

(1) Horst, Paul, "Determination of Optimal Test Length to Maximize the Multiple Correlation," Psychometrika, 1949, 14, pp. 79-88.

(2) Horst, Paul, "A Technique for the Development of a Differential Prediction Battery," Psychological Monographs, 1954, 68, No. 9 (Whole No. 380).

(3) Horst, Paul, Servant of the Human Sciences, Chapter 21, Section 7. Dittoed manuscript. Division of Counseling and Testing Services, University of Washington

## FOOTNOTES

[1] This research was carried out under Contract Nonr-477(08) between the University of Washington and the Office of Naval Research. Most of the computations were carried out by Robert Dear, Charlotte MacEwan and Donald Mills. Much credit is due the typist, Elizabeth Cross. Supervision of both computational and editorial activities was provided by William Clemans. To each of these able contributors I am deeply grateful.

# Armed Services Technical Information Agency

Because of our limited supply, you are requested to return this copy WHEN IT HAS SERVED YOUR PURPOSE so that it may be made available to other requesters. Your cooperation will be appreciated.

# AD 47733

NOTICE: WHEN GOVERNMENT OR OTHER DRAWINGS, SPECIFICATIONS OR OTHER DATA ARE USED FOR ANY PURPOSE OTHER THAN IN CONNECTION WITH A DEFINITELY RELATED GOVERNMENT PROCUREMENT OPERATION, THE U. S. GOVERNMENT THEREBY INCURS NO RESPONSIBILITY, NOR ANY OBLIGATION WHATSOEVER; AND THE FACT THAT THE GOVERNMENT MAY HAVE FORMULATED, FURNISHED, OR IN ANY WAY SUPPLIED THE SAID DRAWINGS, SPECIFICATIONS, OR OTHER DATA IS NOT TO BE REGARDED BY IMPLICATION OR OTHERWISE AS IN ANY MANNER LICENSING THE HOLDER OR ANY OTHER PERSON OR CORPORATION, OR CONVEYING ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE OR SELL ANY PATENTED INVENTION THAT MAY IN ANY WAY BE RELATED THERETO.



# UNCLASSIFIED